Title: A Sensitivity Analysis of a Randomised Controlled Trial of Zinc in Treatment of Falciparum Malaria in Children

Running head: A Sensitivity Analysis of a Randomised Controlled Trial

Matthew P Fox,*† Timothy L Lash, ‡ Davidson H Hamer*¶, #

*Center for International Health and Development, Boston University School of Public Health

‡Department of Epidemiology, Boston University School of Public Health

¶ Division of Geographic Medicine and Infectious Diseases, Tufts-New England Medical Center

Department of Medicine, Tufts University School of Medicine

†Author for correspondence. Address: Center for International Health and Development, Boston University School of Public Health, 85 East Concord St 5th Floor, Boston, MA 02118 USA, mfox@bu.edu, tel: 617-414-1273 fax: 617-414-1261.

Word count: 2240

Sources of support: US Agency for International Development

Abstract

Background: The randomised trial has long been recognized as a paradigm for assessing the efficacy of new interventions, because the investigator can reduce or eliminate many sources of error. As such, clinical trials rarely include quantitative assessments of the extent that systematic error could affect their results. We examined the impact of different sources of bias on a randomised control trial of the efficacy of zinc as an adjuvant to malaria therapy in reducing time to total parasite clearance.

Methods: Using data from a previously published study, we identified two sources of bias and used the sensitivity analysis technique developed by Lash and Fink to assess the impact of each source of bias on the outcome.

Results: After correcting for each source of bias and reincorporating random error into our results, the point estimate of effect comparing those who received placebo to those who received zinc changed slightly (from 0.92 to 0.90) but the 95% interval increased 22% (changing from 0.73 – 1.16 in the conventional analysis to 0.65 – 1.26 in the sensitivity analysis).

Conclusions: The findings of this sensitivity analysis serve as a reminder that the frequentist confidence interval underestimates the total error, even in a randomised control trial. Authors of randomised controlled trial investigations ought to conduct a complete assessment of the impact of potential sources of bias in their studies. CONSORT guidelines for reporting trial results should be updated to encourage authors to assess the impact of non-random errors on their studies.

Keywords: Sensitivity analysis; bias; randomised controlled trials; epidemiologic methods

Abstract Word Count: 244

Introduction

The randomised trial has long been recognized as a paradigm for assessing the efficacy of new interventions, particularly when studying small effects (1). When ethically feasible and when a large enough study can be conducted, a randomised trial is usually the recommended approach for studying new interventions. This design is preferred because the investigator can reduce or eliminate many of the sources of error to which observational studies are prone, such as selection bias, information bias and confounding bias (2). Accordingly, many authors interpret conventional frequentist 95% confidence intervals (CI) derived from rigorously designed randomised trials as if they reflected the total error in the study (3). They assume that by design, systematic error has been reduced to zero, allowing only for random error. As such, clinical trials rarely include quantitative assessments of the extent that systematic error could affect their results. Typical assessments of bias usually focus on whether randomisation "worked" (4), whether the sample size was large enough to detect a difference (5), or whether loss-to-follow up could have explained the results (6-9).

Quantitative sensitivity analysis is a means for assessing the impact of potential biases in studies (10;11). While methods for conducting sensitivity analysis have recently received attention by epidemiologists (12-14), practical software to conduct such analyses is not widely available. Using the sensitivity analysis technique developed by Lash and Fink (12), we have examined the impact of different sources of bias on a randomised control trial of the efficacy of zinc supplementation as an adjuvant to malaria therapy in reducing time to total parasite clearance (15).

Methods

Original data collection and analysis

The methods used to collect data for this randomised, double-blind, placebo controlled trial have been described in the report of the conventional analysis (15). Briefly, children aged six to 60 months with fever ≥37.5 ° C and a parasite count of ≥2000 asexual forms of *Plasmodium falciparum/µl* were randomised to receive either zinc or placebo as an adjuvant to standard antimalarial therapy. The study was a multicenter clinical trial conducted in Ecuador, Ghana, Tanzania, Uganda and Zambia from December 1998 through May 2000. Children who had pneumonia at baseline were excluded from the study. The original study enrolled 1087 children, of whom 542 were randomised to receive an age-dependent daily dose of zinc and 545 children were randomised to placebo. The conventional analysis also examined the proportion of subjects who completely cleared the parasite at 72 hours, which is the outcome that will be investigated in this sensitivity analysis.

Sources of bias assessed

To assess the impact of systematic error we hypothesized that the original study was subject to two potential sources of systematic error. Below we describe each source of bias that was identified and included in the sensitivity analysis in the order they were corrected. Figure 1 shows the steps we took to reconstruct the dataset. Table 1 shows the number of subjects in the conventional analysis stratified by their treatment group and outcome.

(FIGURE 1 HERE)

(TABLE 1 HERE)

Outcome misclassification

We identified two sources of misclassification of the outcome. First, some of the children in the study (n=46 in the zinc arm, n=43 in the placebo arm) did not have a recorded outcome at 72 hours. In the initial analysis, all subjects with no recorded outcome were excluded from the analysis. Some of these subjects likely had a successful outcome and therefore would have been misclassified as to their outcome status. The misclassification is equivalent to missing cases, which, if random, is expected to have no impact on relative measures of effect, but is expected to reduce the precision of the estimates (16). Correcting this bias would therefore be expected to narrow the CI.

To account for the potential impact of the misclassification, we predicted the probability of treatment success for each subject with no recorded outcome. Because treatment success is a function of baseline covariates, we determined the probability of success for each individual with no recorded outcome by fitting a logistic regression model for all those subjects who had a recorded outcome. To account for the random error in the estimated coefficients, we added to each predicted beta coefficient its standard deviation multiplied by a randomly selected standard normal deviate. By repeating this process 2000 times we created a distribution of predictive models. We could then input the actual measured covariates for each subject with no recorded outcome into each model. This process yielded 2000 predicted probabilities of the

outcome per subject who had missing data. We then used this predicted probability to assign an outcome to each person missing an outcome by conducting a Bernoulli trial based on a binomial distribution with a probability equal to the predicted probability from the logistic model. If the Bernoulli trial predicted the subject would have been a treatment success, they were re-classified as to the outcome.

We also identified a second source of outcome misclassification. Despite recent evidence casting doubt on the benefits of use of paracetamol in febrile children (17), many of the subjects received paracetamol for management of fever during the study follow-up period (n=386 in the zinc arm, n = 360 in the placebo arm). While this was not a violation of the study protocol, studies have indicated that, while reducing the amount of time that a fever takes to resolve, use of paracetamol may increase the time to reduction of parasite count (18-21). Accordingly, some of the children who were given paracetamol and who did not clear *P. falciparum* by 72 hours might have done so had they not taken paracetamol. We used the data from those children who did not use paracetamol to predict the probability that those who did use paracetamol would have cleared their parasite by 72 hours had they not taken the paracetamol. We followed the same procedure as above to assign whether each child who received paracetamol was to be reclassified. For each child who received paracetamol we repeated the process 2000 times with a probability distribution about the estimates to determine a range of predicted probabilities.

Because paracetamol use was similar between the two groups, we expected that this would have little impact on the measure of effect, which was already close to the null. However, since there would be greater variability in the predicted outcomes, we

hypothesized that repredicting the outcomes for those with paracetamol use would increase the width of the interval.

Children whose illness was not cause by malaria

While all the children in the study had a high enough parasite count and fever to meet the study definition of acute uncomplicated falciparum malaria, some of the children in the study might have had asymptomatic malaria parasitemia plus fever unrelated to malaria. While this could not have been known upon enrollment, if these children did not have fever due to malaria, then ideally they should have been excluded from the study. We examined all children who went on to develop WHO-defined community-acquired pneumonia (22) during the first 72 hours of follow-up (n=115 in the zinc arm, n=113 in the placebo arm) to assess if they should have been ineligible for the study.

To assign which of these children should not have been in the study, we used estimates from the literature on the probability of presenting with malaria and fever but also having illness unrelated to malaria. While little empirical data exists, we were able to locate some estimates (18;23;24). For each subject who had a recorded diagnosis of respiratory illness we chose the probability that the subject should not have been in the study from a triangular probability distribution, with a minimum probability of 0.086, a maximum probability of 0.85 and a mode of 0.44 (23). Subjects who were assigned to be ineligible for the study were then deleted. Because we used the reconstructed dataset from the last two steps of the reconstruction, each subject had 2000 records, and for each

record we again drew randomly from a triangular probability distribution with the same parameters. This meant that in one iteration a subject may have been removed from the dataset, while in another iteration the subject may have remained.

Because this would have the effect of removing subjects from the dataset, and because the number of children who potentially could be removed from the dataset was similar in each treatment group, we anticipated that this would have minimal effect on the measure of effect, and would increase the width of the confidence interval.

Outcomes reconstructed

We chose to assess the impact of all of the biases described above on complete parasite clearance at 72 hours. Because we corrected each source of bias multiple times, we now had 2000 complete datasets. For each dataset we calculated a standardized morbidity ratio (SMR) and the corresponding standard error. The distribution of these 2000 estimates of the SMR was then used to create simulation intervals that accounted for the source of bias. To account for random error, for each reconstructed dataset, we sampled with replacement (bootstrapping (25)) the same number of subjects as was in the original dataset (1087). For each of the 2000 bootstrapped datasets we calculated an SMR and its estimated standard error to incorporate both random and systematic error.

Results

In the conventional analysis, the Zinc Against Plasmodium Study Group (ZAP), found no changes in (a) time to fever reduction, (b) differences in the proportion of children who had a reduction of 75% or more of their baseline parasite levels at 72 hours after randomisation, or (c) the proportion of children who had total parasite clearance at 72 hours after randomisation (15).

Figure 2 shows the conventional SMR (before accounting for the sources of bias) of total parasite clearance at 72 hours comparing those who received zinc to those who received placebo (SMR=0.92; 95% CI 0.73 – 1.16).

After reconstructing the dataset for all three sources of bias and then reincorporating the effects of random error through bootstrapping, there was a small shift in the point estimate, from SMR=0.92 to SMR=0.90, but a substantial expansion of the confidence interval about the point estimate (from 95% CI of 0.92 – 1.16 to a 95% simulation interval of 0.65 – 1.26). Results of each of the components of the sensitivity analysis can be seen in Table 2. After correcting for both systematic and random error, the width of the confidence interval represents an increased width of 22% over the conventional analysis.

(FIGURE 2 HERE)

As was anticipated, the effect of the sensitivity analysis was to increase the width of the simulation interval, while having little effect on the point estimate. Much of this increase appears to come from the effect of correcting for paracetamol use, which increased the interval 20% when corrected alone.

(TABLE 2 HERE)

Discussion

The results of this analysis give a more complete presentation of the results than do the results from the conventional analysis. The increased width of the confidence interval about the estimate of effect reflects the fact that the study size was reduced by removing subjects who may have been ineligible, which reduced the number of events. From the sensitivity analysis results, we can see that both the corrections for misclassification and the correction of ineligible subjects widened the bootstrap interval. The impact of correcting for paracetamol use also widened the interval once we accounted for random error.

When correcting for both sources of bias simultaneously, the total number of outcome events was reduced, and thereby reduced the precision of the estimates of effect. The fact that the point estimate of the SMR did not change much from the conventional analysis suggests that the effects were comparable in the two treatment groups, most likely because randomisation balanced these external influences on parasite clearance.

The findings of this sensitivity analysis serve as a reminder that the frequentist 95% confidence interval underestimates the total error, even in a randomised control trial. Authors of randomised controlled trial investigations ought to conduct a complete assessment of the impact of potential sources of bias. In this sensitivity analysis we found that the width of the simulation interval derived in the sensitivity analysis was almost 25% larger than the conventional frequentist interval.

The CONSORT statement on the reporting of clinical trials states that authors should include in their discussion section, "Interpretation of the results, taking into account study hypotheses, *sources of potential bias or imprecision* and the dangers

associated with multiplicity of analyses and outcomes." (26) (italics ours). Whether or not authors include all of the potential sources of bias in randomised controlled trials is a matter of speculation; however, the CONSORT statement does not recommend that authors quantify the sources of potential bias, even by making assumptions about the direction and magnitude of these biases or through incorporating the bias into their estimates of effect through some of the methods described by Greenland (10).

This sensitivity analysis demonstrates that the usually relied upon frequentist confidence interval, even from a well designed randomised trial, cannot be assumed to include the total error of a study. Even when authors give statements about the possible direction and magnitude of possible biases in their discussion sections, it is difficult for consumers of the literature to estimate the impact that multiple sources of bias will have when acting simultaneously. Studies that incorporate sensitivity analysis can include the systematic error in their measures of effect and confidence interval and can reduce the burden on the reader without increasing the amount of space necessary to present results.

Figure 1 – Sensitivity Analysis Flowchart

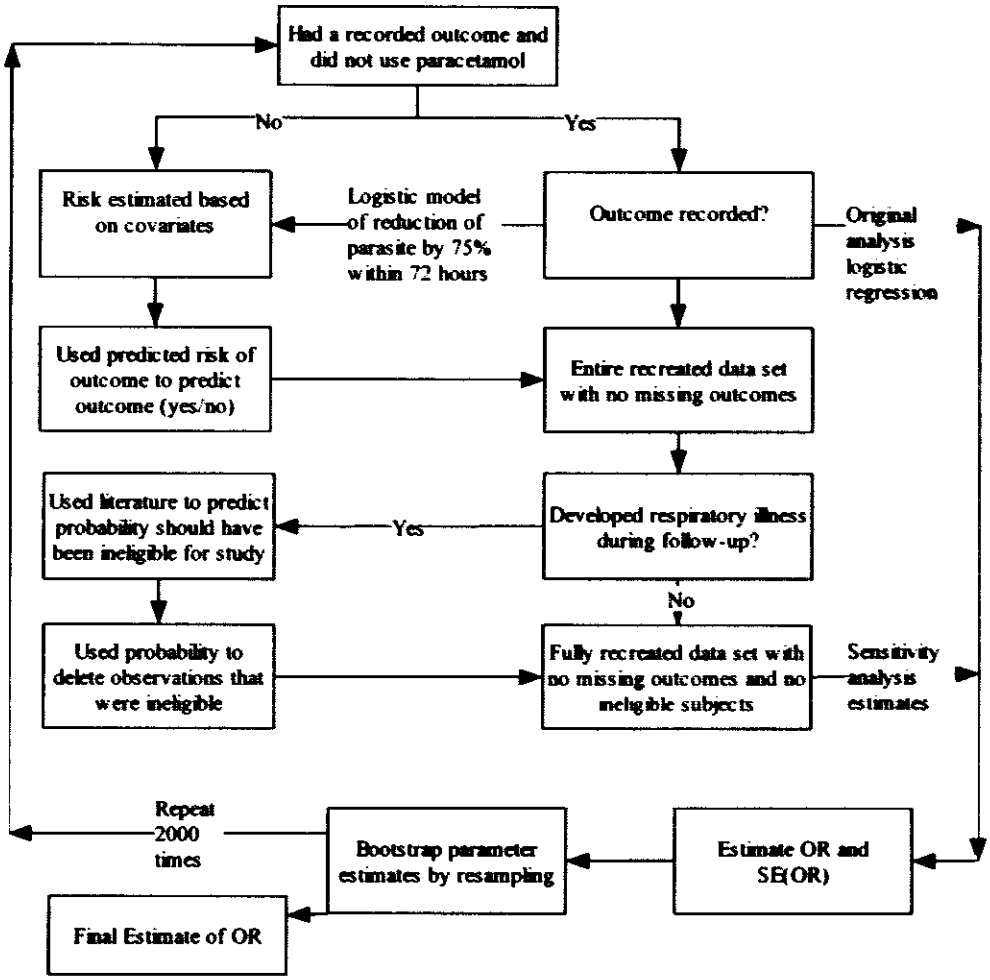

Figure 2 – Sensitivity Analysis of a Trial of Zinc as an Adjuvant to Malaria Therapy

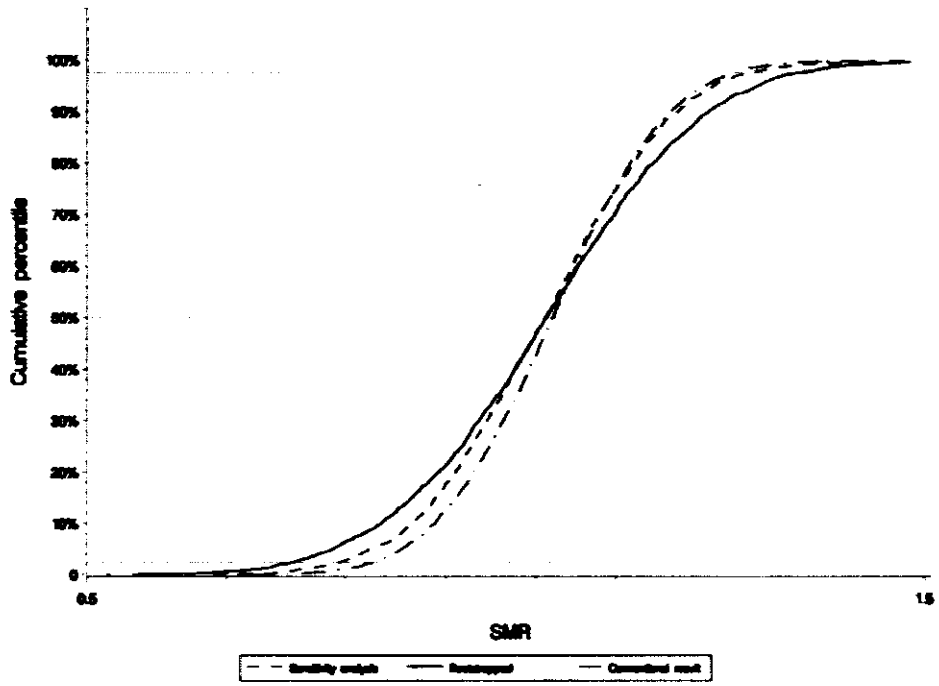

Table 1 - Number of Subjects by Treatment Group, Outcome, and Paracetamol Use

Number of Subjects by Treatment Group, Outcome, and Paracetamol Use

	No Paracetamol			Paracetamol			Total		
	Placebo	Zinc	Total	Placebo	Zinc	Total	Placebo	Zinc	Total
Treatment Success	59	46	105	49	54	103	108	100	208
Treatment Failure	98	90	188	296	306	602	394	396	790
No Outcome	28	20	48	15	26	41	43	46	89
Total	185	156		360	386		545	542	1087

Table 2 – Sensitivity Analysis of a Trial of Zinc as an Adjuvant to Malaria Therapy

Systematic Error [1]

	2.5%	50%	97.5%	Width CI
Conventional analysis	0.92	0.92	0.92	1.00
Sensitivity analysis correcting for:				
Those with no recorded outcome	0.85	0.90	0.96	1.14
Those who took Paracetamol	0.69	0.91	1.19	1.72
Those with illness not caused by malaria	0.86	0.88	0.92	1.07
All three simultaneously	0.70	0.90	1.18	1.69

Systematic and Random Error [1]

	2.5%	50%	97.5%	Width CI	% Increase *
Conventional analysis (random error only)	0.73	0.92	1.16	1.59	NA
Sensitivity analysis correcting for:					
Those with no recorded outcome	0.71	0.90	1.16	1.64	3%
Those who took Paracetamol	0.66	0.91	1.26	1.91	20%
Those with illness not caused by malaria	0.66	0.89	1.15	1.73	8%
Bootstrap (random and systematic error)	0.65	0.90	1.26	1.95	22%

* The % increase compares each analysis to the conventional analysis accounting for random error only

[1] All analyses use the placebo group as the reference group

Acknowledgements

Major funding for the research presented in this paper was provided by US Agency for International Development through the Child Health Research Project, G/PHN/HN/CS, Global Bureau, USAID, under the terms of Cooperative Agreement No HRN-A-00-96-90010-00, the Applied Research on Child Health (ARCH) Project. The opinions expressed herein are those of the authors and do not necessarily reflect the views of the US Agency for International Development.

Reference List

(1) MacMahon S, Collins R. Reliable assessment of the effects of treatment on mortality and major morbidity, II: observational studies. *Lancet* 2001; **357**:455-462.

(2) Rothman K, Greenland S. Precision and validity in epidemiologic studies. In: Rothman KJ, Greenland S, (eds). *Modern Epidemiology, 2nd Edn*. Philadelphia, PA: Lippincott-Raven, 1998: 115-134.

(3) Brennan P, Croft P. Interpreting the results of observational research: chance is not such a fine thing. *BMJ* 1994; **309**:727-730.

(4) Altman DG. Randomisation and baseline comparisons in clinical trials. *Lancet* 1990; **335**:149-153.

(5) Altman DG, Bland JM. Absence of evidence is not evidence of absence. *BMJ* 1995; **311**:485.

(6) Touloumi G, Pocock SJ, Babiker AG, Darbyshire JH. Impact of missing data due to selective dropouts in cohort studies and clinical trials. *Epidemiology* 2002; **13**:347-355.

(7) Ko CY, Sack J, Chang JT, Fink A. Reporting randomized, controlled trials: where quality of reporting may be improved. *Dis Colon Rectum* 2002; **45**:443-447.

(8) DerSimonian R, Charette LJ, McPeek B, Mosteller F. Reporting on methods in clinical trials. *N Engl J Med* 1982; **306**:1332-1337.

(9) Emerson JD, McPeek B, Mosteller F. Reporting clinical trials in general surgical journals. *Surgery* 1984; **95**:572-579.

(10) Greenland S. Basic methods for sensitivity analysis of biases. *Int J Epidemiol* 1996; **25**:1107-1116.

(11) Greenland S. Basic methods for sensitivity analysis and external adjustment. In: Rothman KJ, Greenland S, (eds). *Modern Epidemiology, 2nd Edn*. Philadelphia, PA: Lippincott-Raven, 1998: 343-358.

(12) Lash T, Fink AK. Semi-automated sensitivity analysis to assess systematic errors in observational data. *Epidemiology* 2003; **14**:451-458.

(13) Greenland S. The impact of prior distributions for uncontrolled confounding and response bias. *J Am Stat Assoc* 2003; **98**: 47-54.

(14) Phillips CV. Quantifying and reporting uncertainty from systematic errors. *Epidemiology* 2003; **14**:459-466.

(15) Zinc Against Plasmodium Study Group. Effect of zinc on the treatment of Plasmodium falciparum malaria in children: a randomized controlled trial. *Am J Clin Nutr* 2002; **76**:805-812.

(16) Brenner H, Savitz DA. The effects of sensitivity and specificity of case selection on validity, sample size, precision, and power in hospital-based case-control studies. *Am J Epidemiol* 1990; **132**:181-192.

(17) Russell FM, Shann F, Curtis N, Mulholland K. Evidence on the use of paracetamol in febrile children. *Bull World Health Organ* 2003; **81**:367-372.

(18) Brandts CH, Ndjave M, Graninger W, Kremsner PG. Effect of paracetamol on parasite clearance time in Plasmodium falciparum malaria. *Lancet* 1997; **350**:704-709.

(19) Lell B, Sovric M, Schmid D, Luckner D, Herbich K, Long HY et al. Effect of antipyretic drugs in children with malaria. *Clin Infect Dis* 2001; **32**:838-841.

(20) Kazembe PN. Effect of paracetamol on parasite clearance time in Plasmodium falciparum malaria. *Lancet* 1997; **350**:1776.

(21) Kluger MJ, Kozak W, Conn CA, Leon LR, Soszynski D. The adaptive value of fever. Infect *Dis Clin North Am* 1996; **10**:1-20.

(22) World Health Organization. *Technical Bases for the WHO Recommendations on the Management of Pneumonia in Children at First-Level Facilities*. WHO/ARI/91.20. Geneva: World Health Organization, 1991.

(23) World Health Organization. *The Overlap in the Clinical Presentation and Treatment of Malaria and Pneumonia in Children: Report of a Meeting*. WHO/MAL/92.1065 . Geneva: World Health Organization, 1991.

(24) Byass P, Campbell H, O'Dempsey TJ, Greenwood BM. Coincidence of malaria parasitaemia and abnormal chest X-ray findings in young Gambian children. *J Trop Med Hyg* 1991; **94**:22-23.

(25) Carpenter J, Bithell J. Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *Stat Med* 2000; **19**:1141-1164.

(26) Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* 2001; **357**:1191-1194.